JPL, AI, and Data Science

Dan Crichton, Leader, Center for Data Science and Technology;
Program Manager, Data Science

Jet Propulsion Laboratory
California Institute of Technology

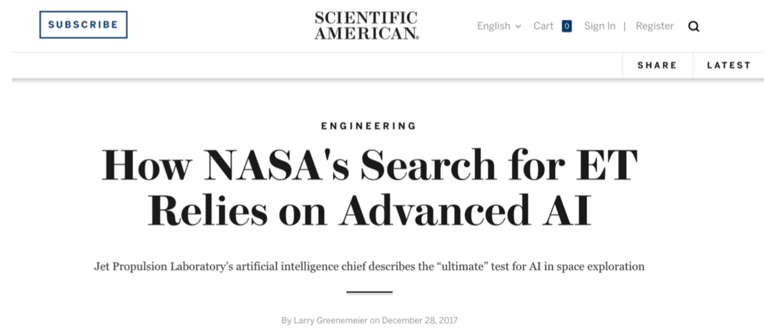
Overview

- JPL Data Science / AI Strategy
- AI Onboard
- Shift toward Data Analytics
- Cross-Cutting Examples
- Partnering

JPL Data Science / AI Strategy

Tackling the AI and Data Challenges at JPL

- JPL is engaging data science and AI technologies and methodologies for science, mission operations, engineering applications.
  - From onboard computing to scalable archives to analytics
  - Applying ML techniques with supporting infrastructure
- JPL has established a program focused on building and implementing an institution-wide strategy for data science and AI
  - Expanding from archives to enable data analytics as a first class activity
  - Methodology transfer across disciplines
  - Research partnerships with academia, government, and industry


SUBSCRIBE SCIENTIFIC AMERICAN English Cart 0 Sign In | Register
SHARE LATEST

ENGINEERING

How NASA's Search for ET Relies on Advanced AI

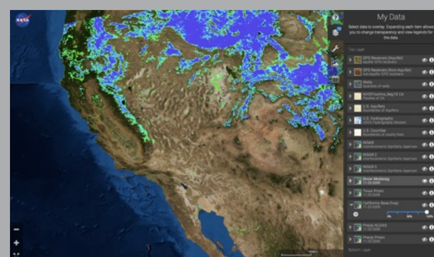
Jet Propulsion Laboratory's artificial intelligence chief describes the "ultimate" test for AI in space exploration

By Larry Greenemeier on December 28, 2017

Driving AI and Data Science into JPL Activities

- 25 pilots launched 2017-18
  - Spanning science, mission and DSN operations, and formulation
  - Building towards a data science vision of full utilization of data and agile application of analytics

Use Cases: Science



Use Cases: Mission Ops

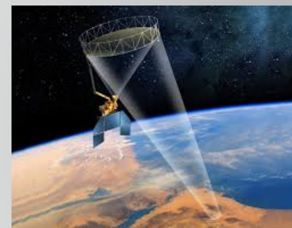
Use Cases: Formulation

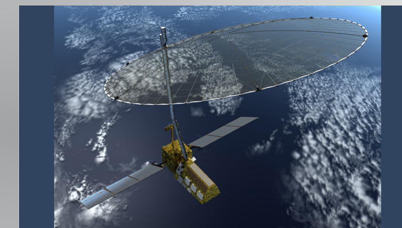
Use Cases: Institution

Applying AI Across the Mission-Science Data Lifecycle

Emerging Solutions
- *Onboard Data Analytics*
- *Onboard Data Prioritization*
- *Flight Computing*

Observational Platforms and Flight Computing

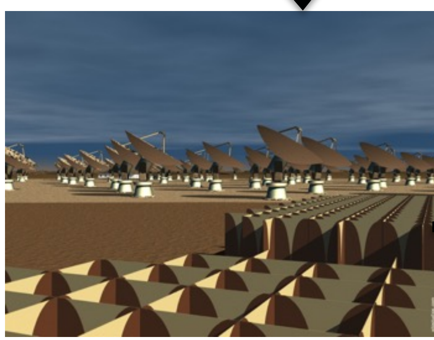




SMAP (Today): 485 GB/day **NI-SAR (2020): 86 TB/day**

(1) Too much data, too fast; cannot transport data efficiently enough to store

Emerging Solutions
- *Intelligent Ground Stations*
- *Agile MOS-GDS*



(2) Data collection capacity at the instrument continually outstrips data transport (downlink) capacity

Ground-based Mission Systems

Massive Data Archives and Big Data Analytics

Emerging Solutions
- *Data Discovery from Archives*
- *Distributed Data Analytics*
- *Advanced Data Science Methods*
- *Scalable Computation and Storage*

(3) Data distributed in massive archives; many different types of measurements and observations

Concept included in OCT TA-11 Roadmap (2015): 11.4.1 – Mission, Science, Engineering Data Lifecycle

AI Onboard

EO-1 (2004): Autonomous Spacecraft AI

The onboard software enabled the spacecraft to detect and track volcanism, flooding, and cryosphere

Increasing Computing Capability Onboard

Heading Toward Multicore in Space

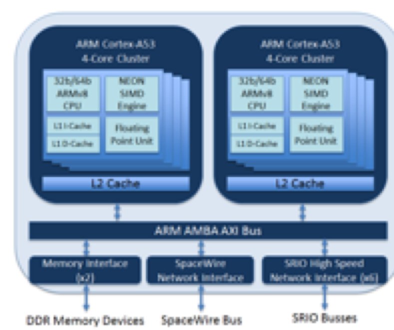
Voyager computer

- 8,000 instructions/sec and kilobytes of memory

iPhone

- 14 GOPS and gigabytes of memory

Curiosity (Mars Science Laboratory)
Processor: 200 MOPS BAE RAD750



HPSC (NASA STMD / USAF)
Processor: 15 GOPS, extensible

Onboard Analysis

Dust Devils on Mars

Dust devils are scientific phenomena of a transient nature that occur on Mars

- They occur year-round, with seasonally variable frequency
- They are challenging to reliably capture in images due to their dynamic nature
- Scientists accepted for decades that such phenomena could not be studied in real-time

Spirit Sol 543 (July 13, 2005)

New onboard Mars rover capability (as of 2006)

- Collect images more frequently, analyze onboard to detect events, and only downlink images containing events of interest

Benefit

- < 100% accuracy can dramatically increase science event data returned to Earth
- *First notification includes a complete data product*

Credit: T. Estlin. B. Bornestein, A. Castano, J. Biesiadecki, L. Neakrase, P. Whelley, R. Greeley, M. Lemmon, R. Castano, S. Chien and MER project team

Surface Mobility

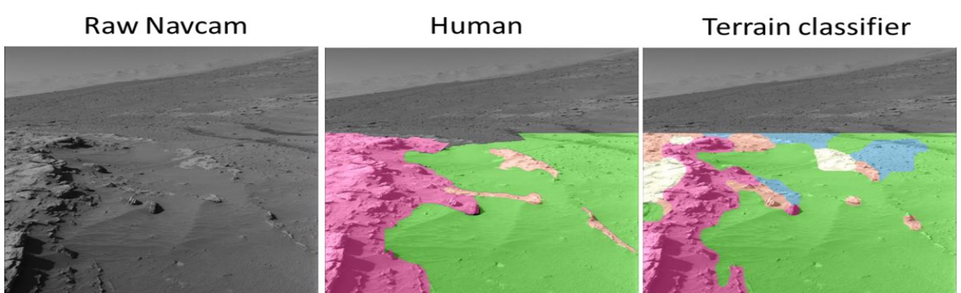
Mars Rover Navigation

Flight Deployed

- **1996 Mars Pathfinder:** obstacle avoidance with structured light
- **2003 Mars Exploration Rover:** obstacle avoidance with stereo vision; pose estimation and slip detection with visual odometry; goal tracking
- **2011 Mars Science Laboratory:** enhanced obstacle avoidance, visual odometry and goal tracking

Research and Development

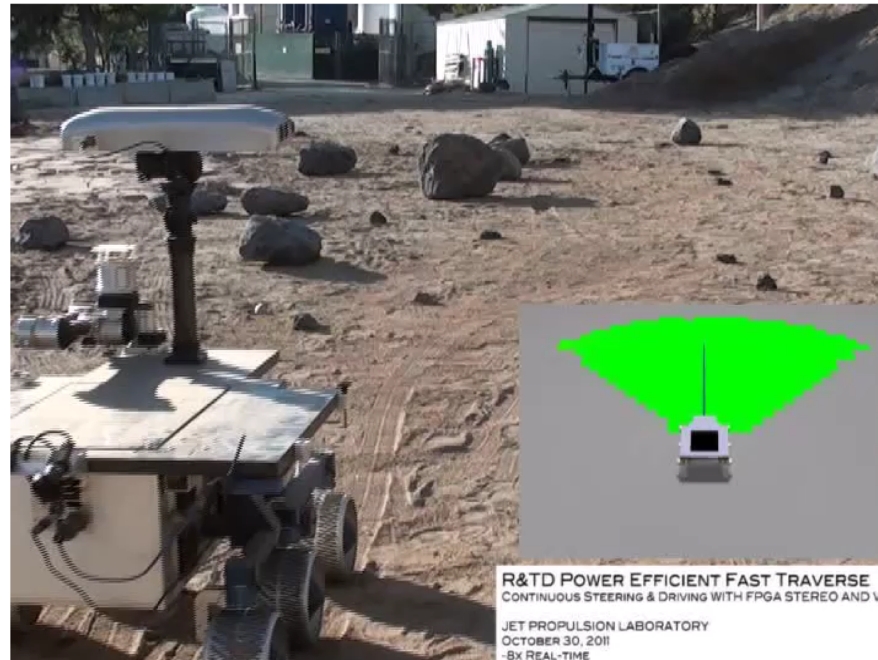
– Enhanced hazard detection, traversability analysis and motion planning for Mars 2020 and beyond



T.Estlin, et al

Terrain Classification

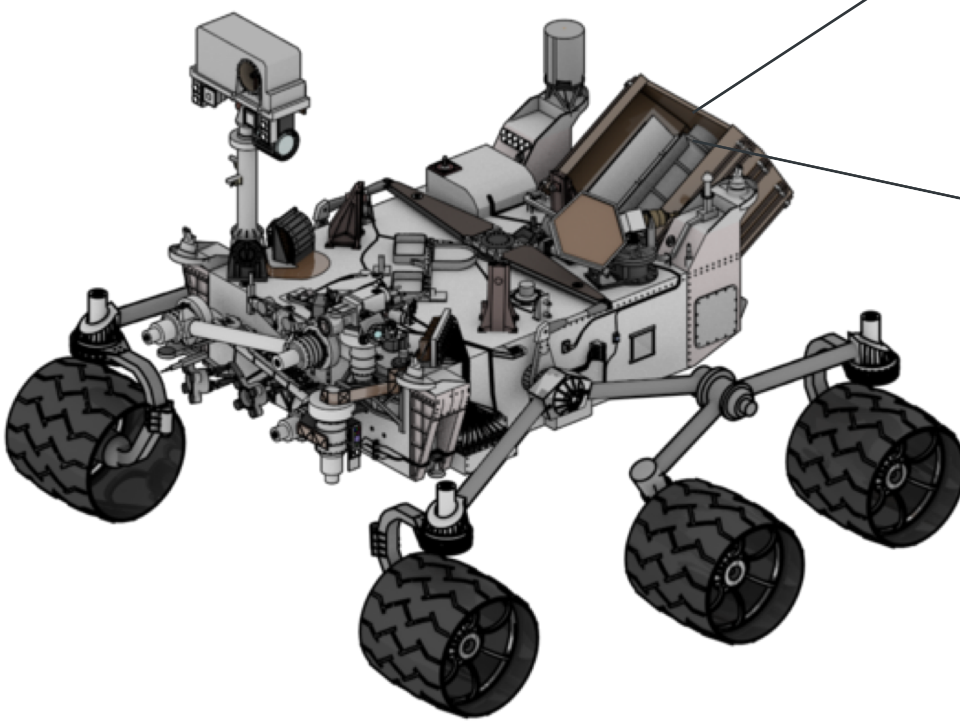
Athena

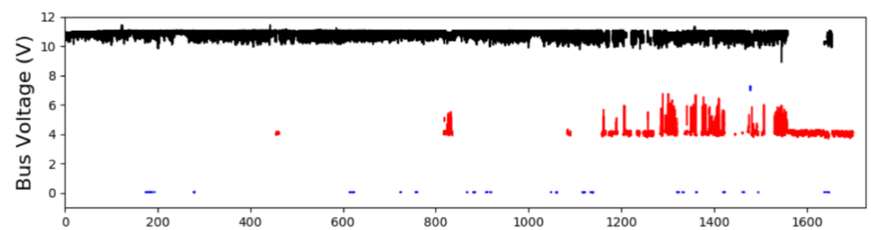


Fido

Onboard Numeric Watchdog for Analysis of Telemetry Channel Heuristics (ON-WATCH)

Assessing Vehicle Anomalies

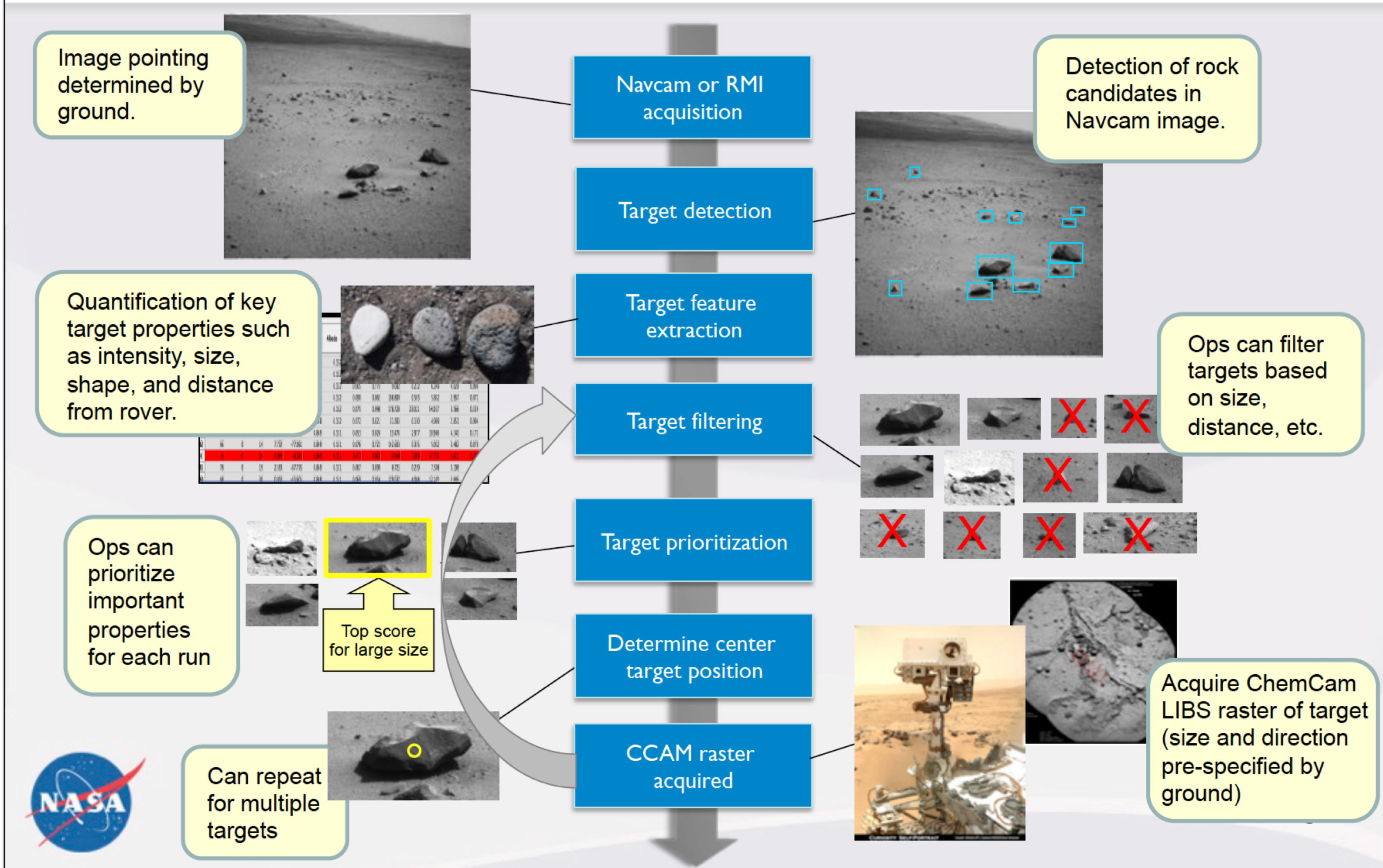




Managing Bandwidth: Detect anomalous behavior using ML techniques for investigation

Jack Lightholder & Lukas Mandrake

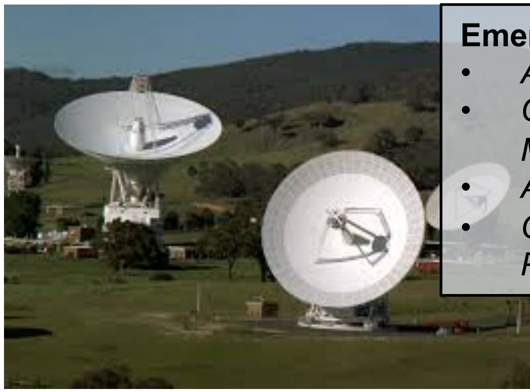
Onboard Process for MSL



Shift toward Data Analytics

Data-Driven Capabilities Across the Ground Environment

Intelligent Ground Stations



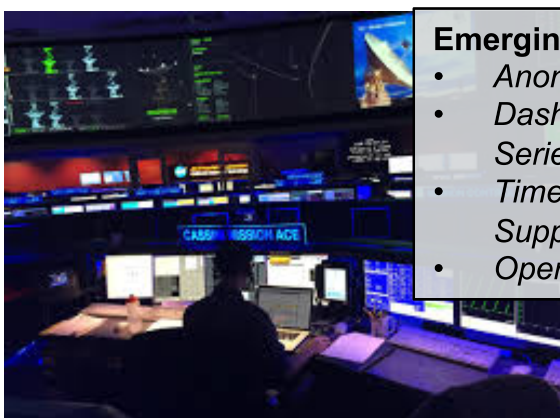
Emerging Solutions
- *Anomaly Detection*
- *Combining DSN & Mission Data*
- *Attention Focusing*
- *Controlling False Positives*

Data-Driven Discovery from Archives

Emerging Solutions
- *Automated Machine Learning - Feature Extraction*
- *Intelligent Search*
- *Integration of disparate data*

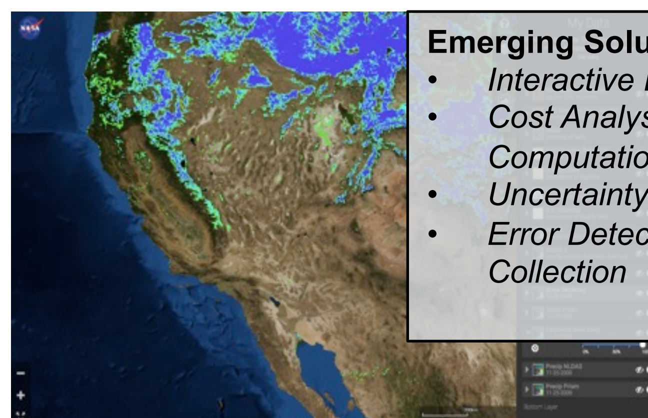
Technologies: Machine Learning, Deep Learning, Intelligent Search, Data Fusion, Interactive Visualization and Analytics

Agile Mission Operations



Emerging Solutions
- *Anomaly Interpretation*
- *Dashboard for Time Series Data*
- *Time-Scalable Decision Support*
- *Operator Training*

Data Analytics and Decision Support



Emerging Solutions
- *Interactive Data Analytics*
- *Cost Analysis of Computation*
- *Uncertainty Quantification*
- *Error Detection in Data Collection*

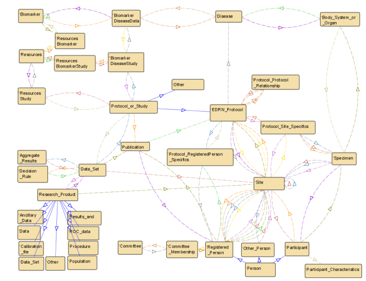
Technical Capabilities Enabling AI & Analytics

R
amazon webservices™
Apache SOFTWARE FOUNDATION
Spark Lightning-Fast Cluster Computing
accumulo™

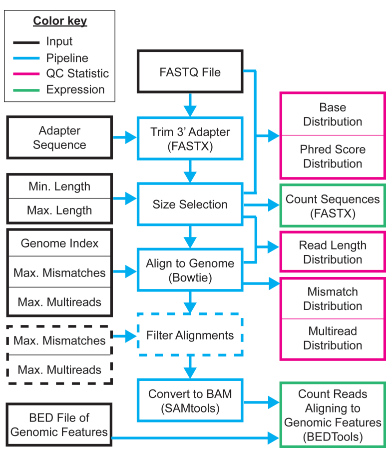
Cloud, Open Source, and Big Data Infrastructures

PYTORCH
mxnet
TensorFlow
Caffe

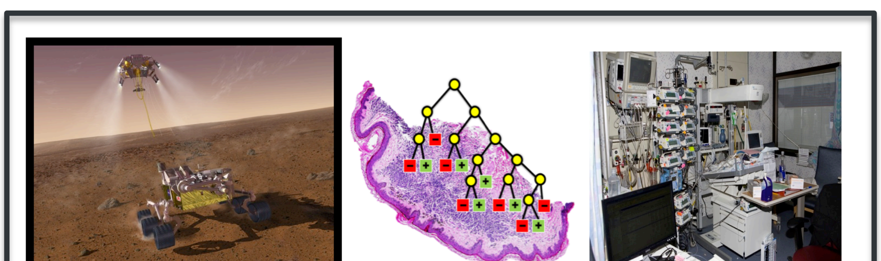
Machine Learning and Deep Learning



Ontologies and Information Models



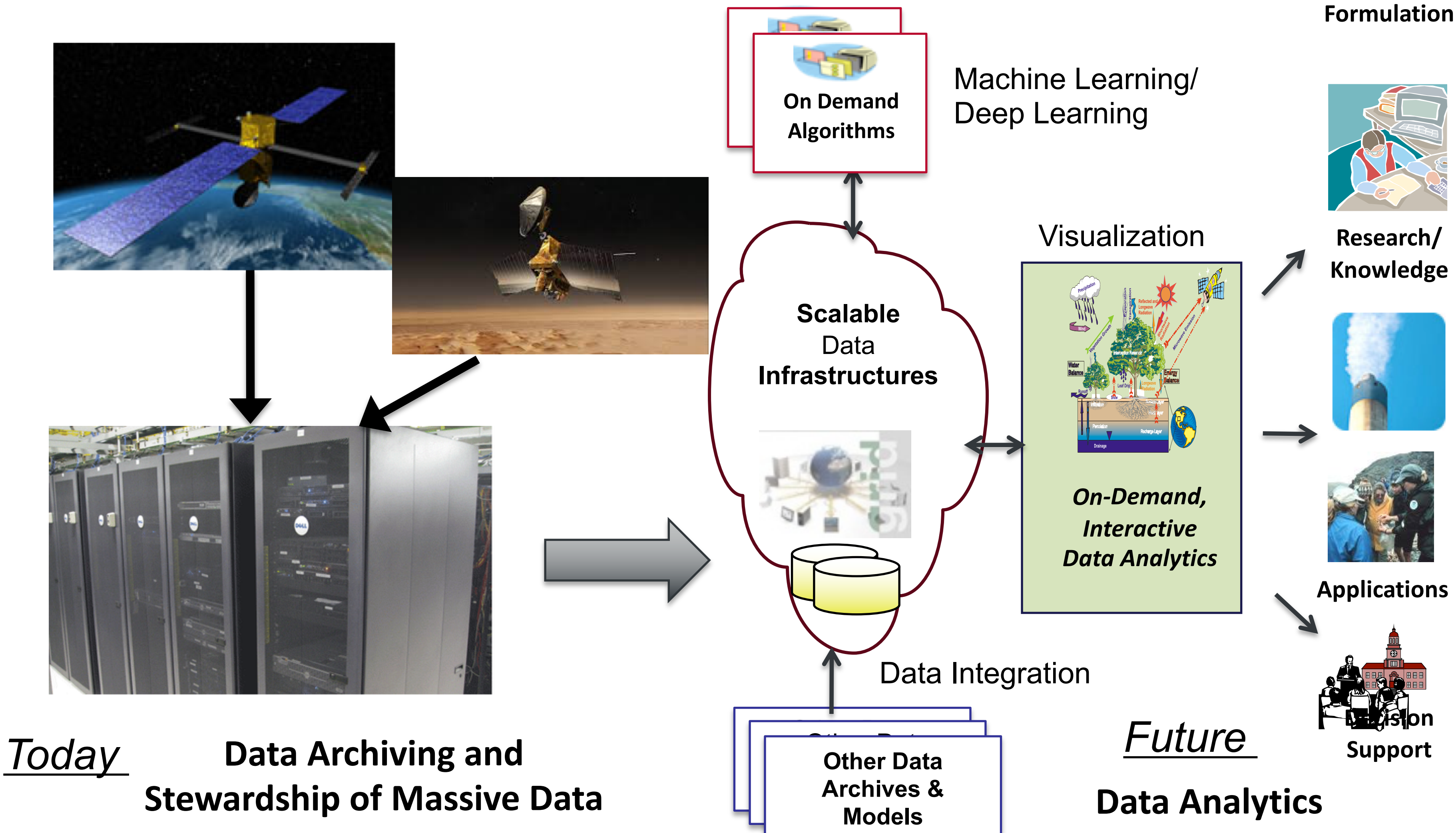
Computational Pipelines/HPC


Planetary
Biology
Healthcare

<u>Great Opportunities for Methodology Transfer and Collaboration</u>

Visualization and HCI Techniques

Expanding to Data-Driven Analytics



Reducing Data Wrangling: "There is a major need for the development of software components… that link high-level data analysis-specifications with low-level distributed systems architectures."
Frontiers in the Analysis of Massive Data, National Research Council, 2013.

Cross-Cutting Examples

Data-Driven Approaches for Deep Space Communication: Detecting Anomalies

Current Inputs: DSN operationally relevant data

Real Time Monitor Data

Real Time Operator Logs

Track Predicts

DSN Software Quality Assessment (SQA) Data Archive

- Relational database
- 10 years of data
- 1.3+ billion records

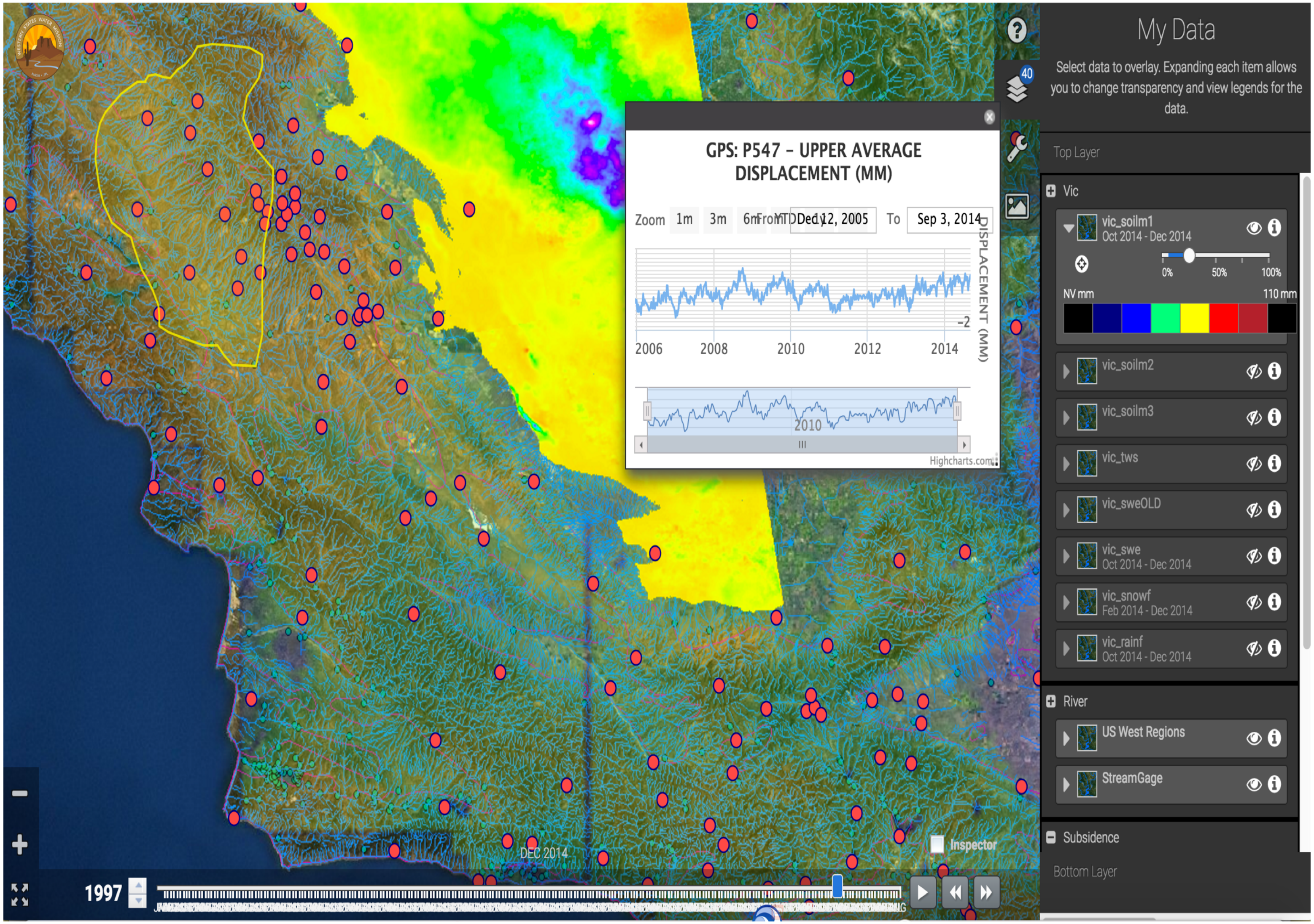
Desired Output: Better Fault Detection and Diagnosis

Real-Time, Historically Informed Alerts

Real-Time Insight into Data Points' Criticality and Relationships

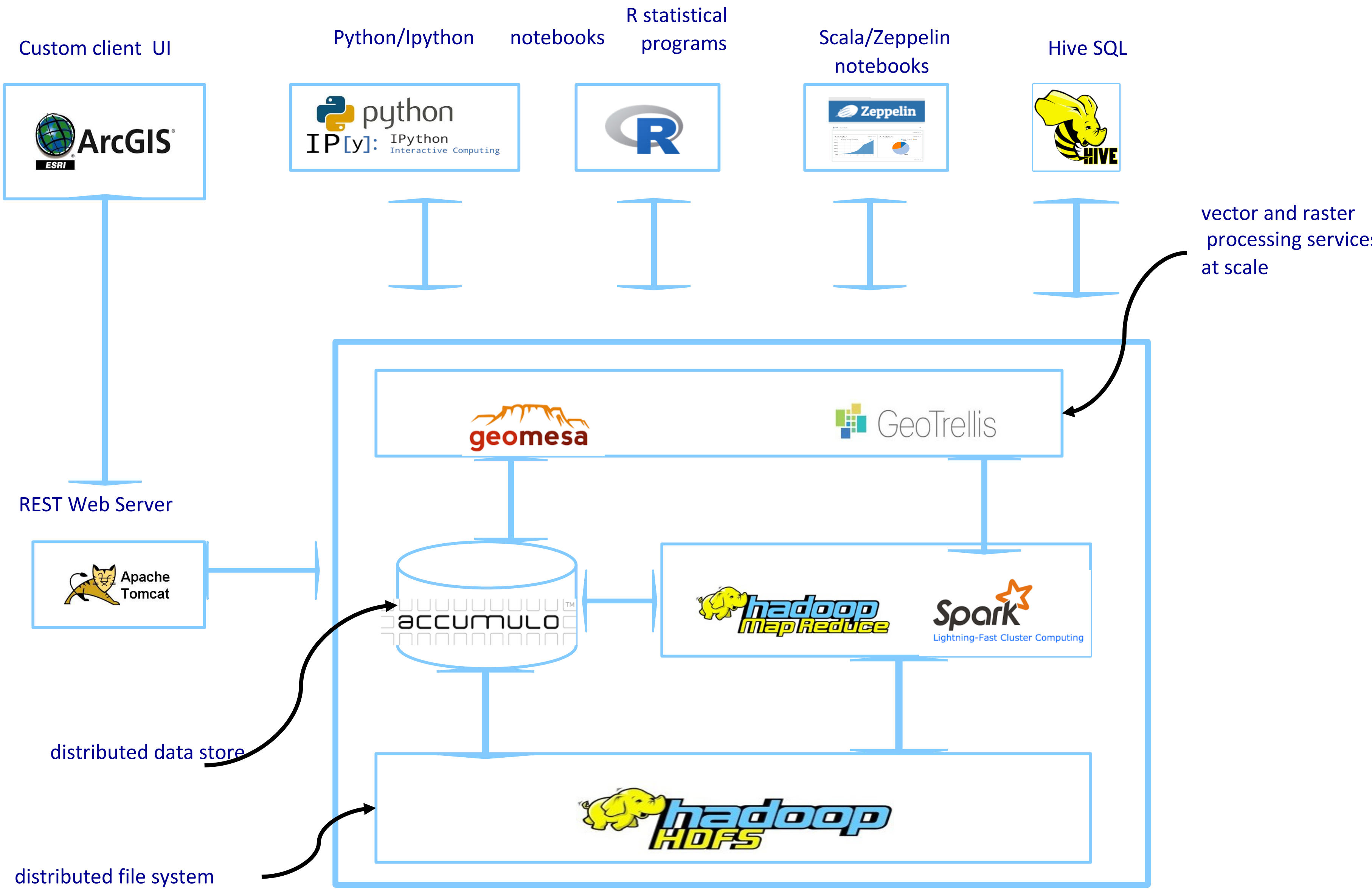
Credit: Rishi Verma, JPL

WaterTrek:
Interactive Analytics for Western States Water Analysis



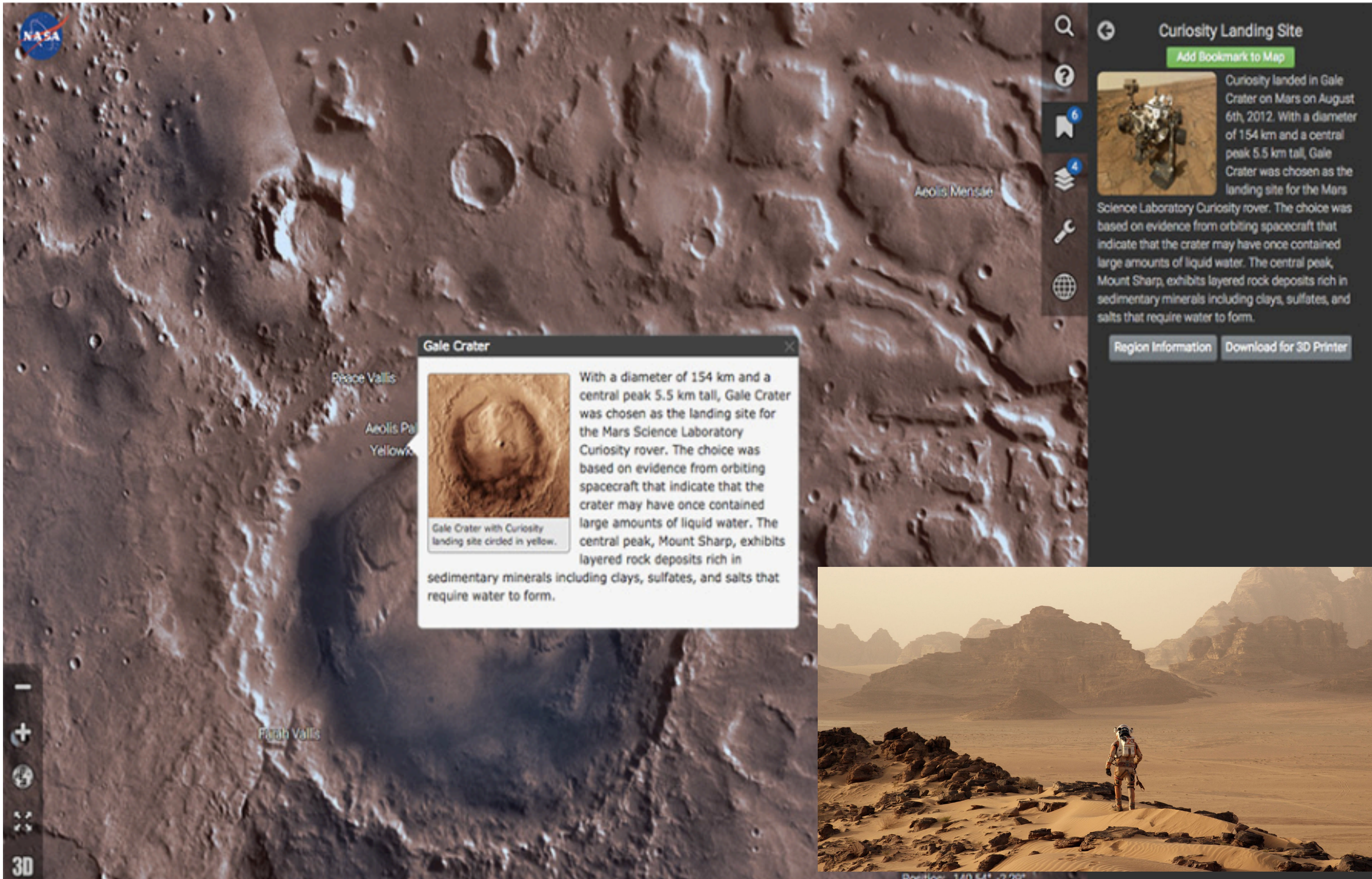
Credit: Jay Famiglietti, Cedric David, Shan Malhotra, D. Crichton

WaterTrek: Analytic Data Infrastructure



Open source and scalable to cloud; 180 billion data points accessible < 1 second

Mars Trek: Interactive Analytics for Exploring Mars

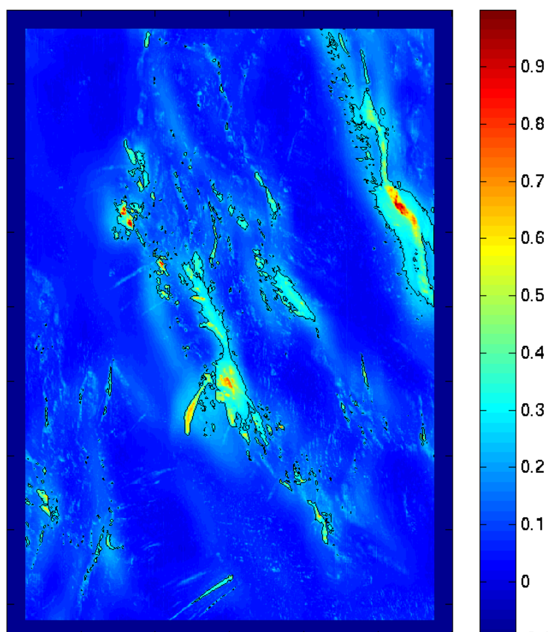


Credit: Emily Law, Shan Malhotra

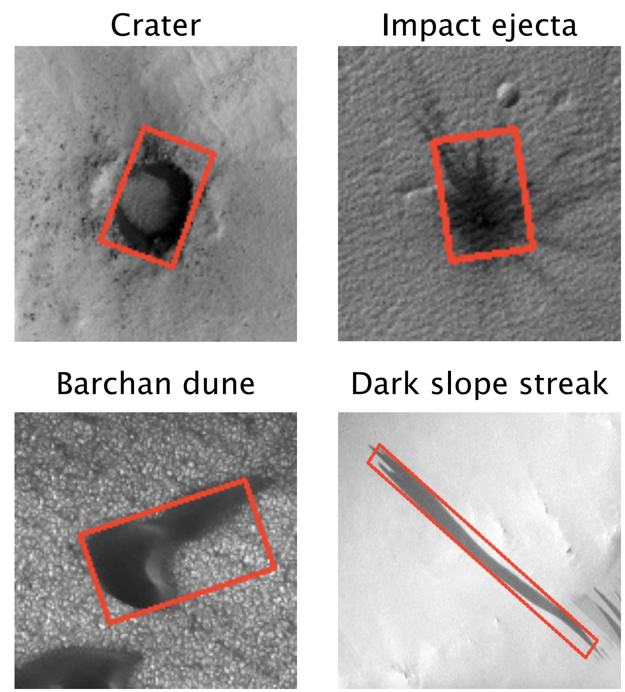
Mars Image Classification

- About ~1.3M images from MRO Mission HiRISE instrument
- Previously no way to easily find images with certain landmarks (e.g., craters)
- New Approach:
    1) Determine high salience (i.e., distinctive) regions by computing statistical differences between pixel and surrounding context
    2) Classify landmarks using *machine learning model* and user *training data*

MOC, June 2000



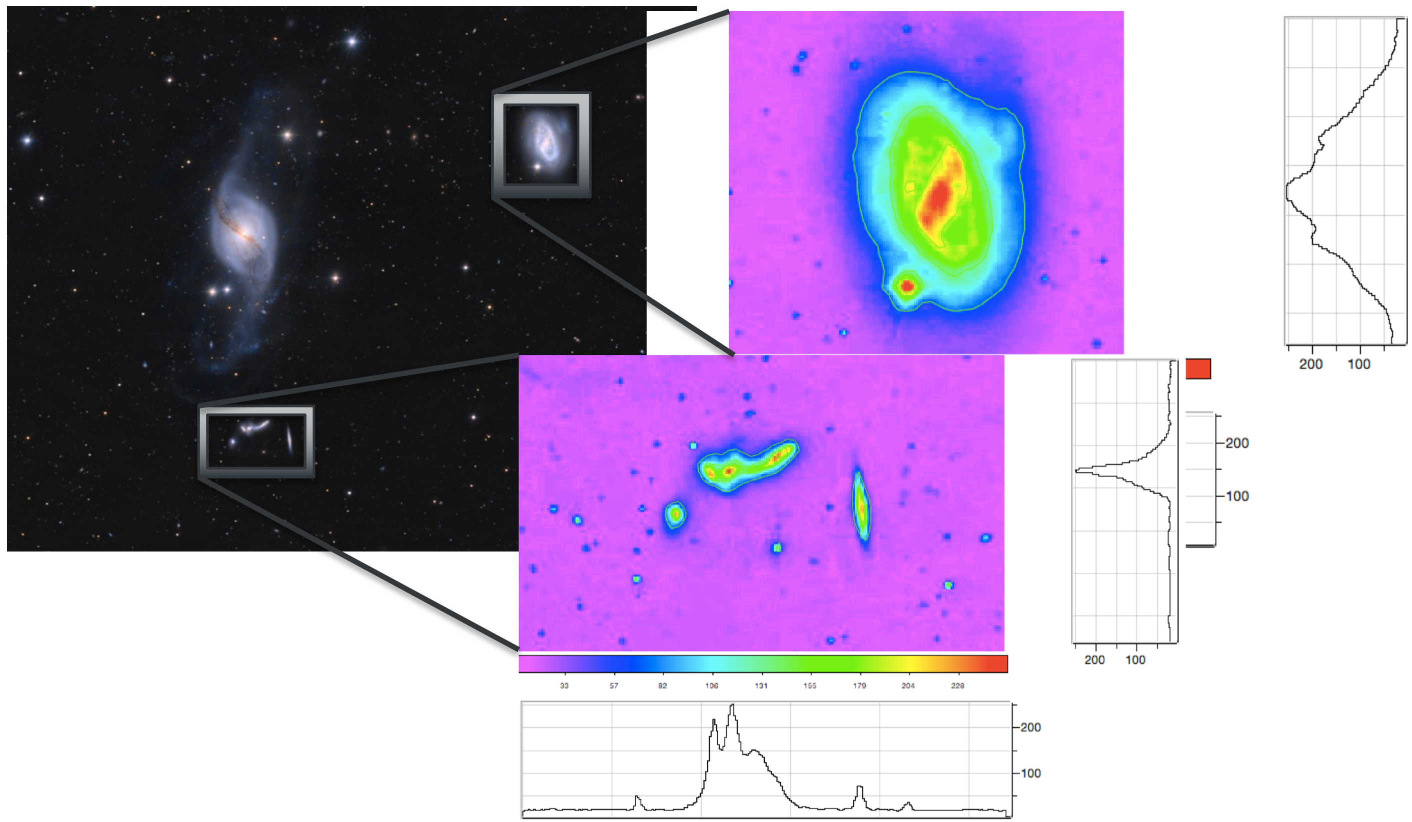
Salience Map



Examples of classified landmarks

POC: K. Wagstaff

Feature Identification in Astronomy Imaging



Description: Detecting objects from astronomical measurements by evaluating light measurements in pixels using machine learning.

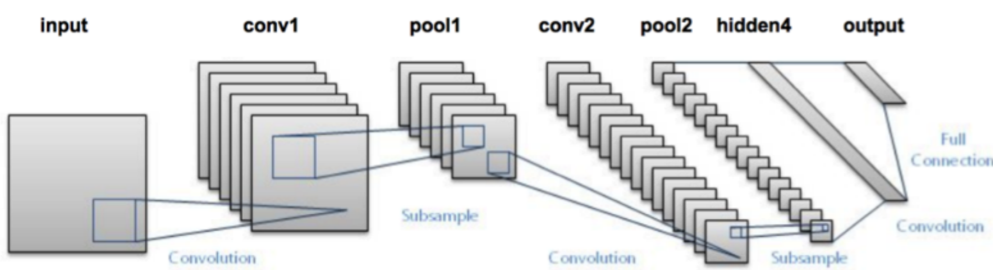
Image Credit: Catalina Sky Survey (CSS), of the Lunar and Planetary Laboratory, University of Arizona, and Catalina Realtime Transient Survey (CRTS), Center for Data-Driven Discovery, Caltech.

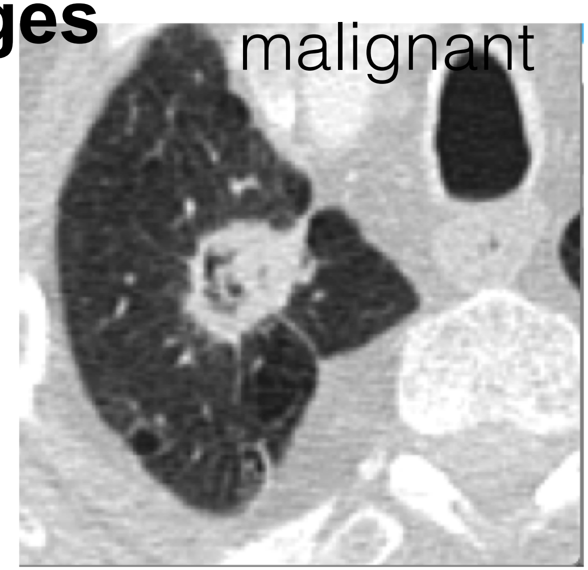
Partnering

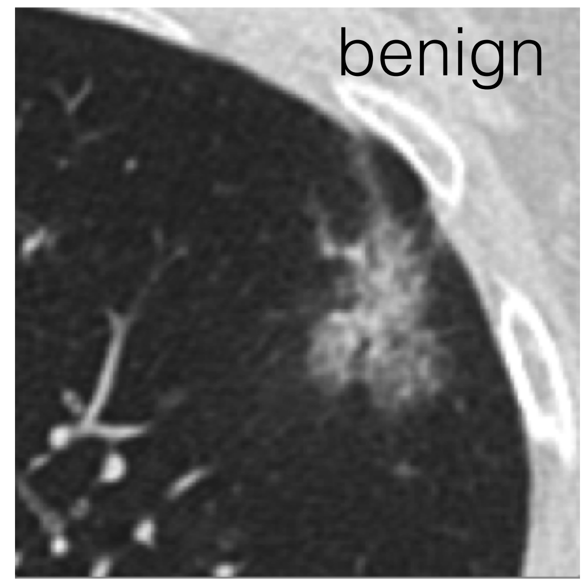
Methodology Transfer: ML and Crowdsourcing to classify features in cancer images

Traditional Machine Learning Flow

Deep Learning Flow







Promise:
Works better

Pitfall:
Blacker box

A. Mahabal, D. Liu, D. Crichton

Caltech-JPL
Partnership in Data Science and AI

Center for Data-Driven Discovery on campus/Center for Data Science and Technology at JPL

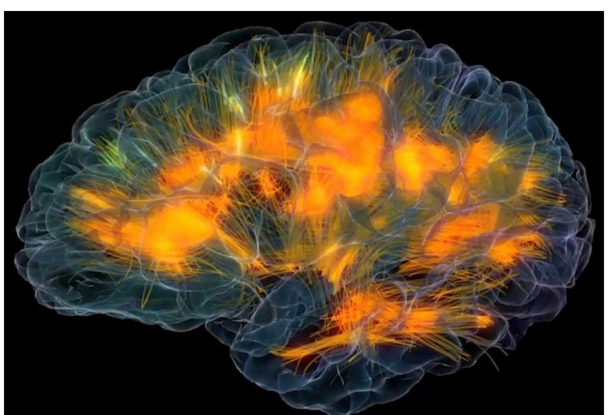
From basic research to deployed systems ~10 collaborations

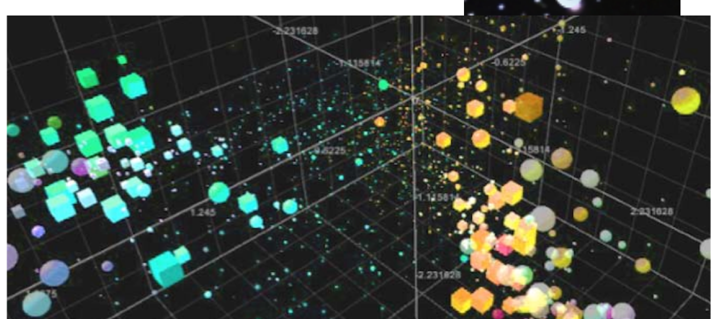
Leveraged funding from JPL to Caltech; from Caltech to JPL

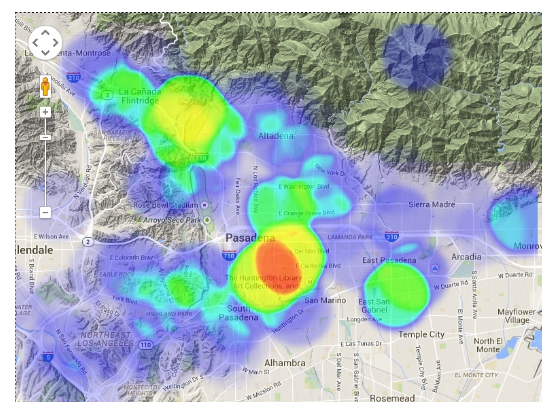
Virtual Summer School (2014) has seen over 25,000 students

CENTER FOR DATA-DRIVEN DISCOVERY

CDST

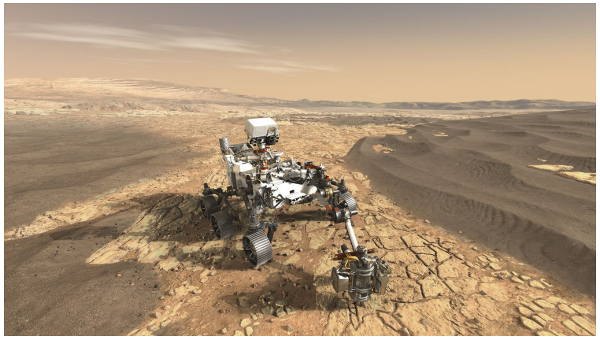




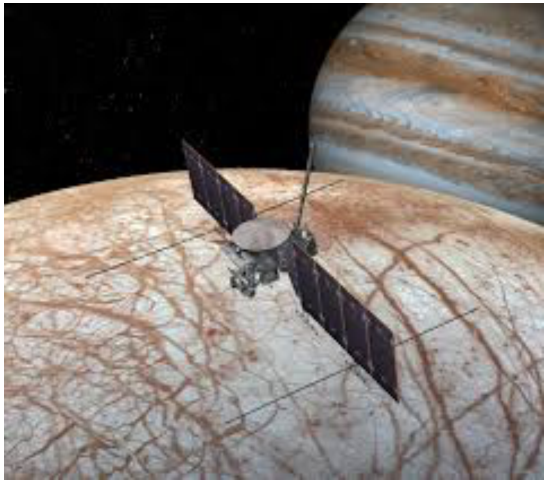


Conclusions

- JPL Strategy is to drive AI and Data Science into the fabric of JPL by
  - Launching cross-institution pilots
  - Building a trained workforce
  - Linking to the mission-science data lifecycle



- Great opportunities to both innovate onboard and leverage emerging capabilities and platforms on the ground
  - Transform autonomy onboard
  - Transform mission operations
  - Drive new science insights



- AI and Data Science will be an essential part of NASA's future!

NASA JPL Caltech